

# 子どもの保護のための Web アクセス監視支援システム

## Web access watchdog system for child's protection

上田 達巳  
Tatsumi Ueda

高井 昌彰†  
Yoshiaki Takai

### 1. はじめに

近年インターネットの利用は極めて一般的なものとなっており、子どもたちもまたインターネットから様々な情報を取得する機会が増えている[1]. インターネットでは、有益な情報を得ることができる反面、リンクを辿っていくことによる意図しない有害な情報との接触や、操作ミスによる個人情報の誤発信、掲示板などで名誉毀損的な発言をしてしまい收拾がつかなくなるといった問題に遭遇してしまう危険性がある. これらの問題から子どもたちを保護するために、インターネット上に存在する有害なコンテンツを排除しようという動きがあるが、現段階では法的に情報発信を規制するには限界がある[2].

このため、受信者側で受け取る情報を選択または制限する方法としてフィルタリングが用いられている[3]. しかし、フィルタリングには過度に情報を制限してしまい、本来有用となる情報を得られない場合があるという問題がある.

また、子どもがインターネットに接続する場合は、操作をしている間、保護者がそばにいて監視することが推奨されているが、常に監視し続けることは難しく、子どもがインターネットで何をしているかを十分に把握しきれていない場合もある.

本稿では、子どものプライバシーに配慮しながらインターネット利用状況を保護者に伝え有害な情報から保護するための Web アクセス監視支援システムについて述べる. また、Windows でのプロトタイプの実装と動作実験について述べる.

### 2. システムの概要

#### 2.1 全体構成

本システムの概略図を図 1 に示す. 本システムは子どもが直接利用する PC 内に設置し、インターネットと PC 間で送受信されるパケットに含まれる Web ドキュメントを監視対象とする. Web ドキュメント内に有害な情報が含まれていないかどうか検知し、電子メールを用いて分析結果の要旨を保護者へと通知する. なお、システムの運用にあたっては、子どもと保護者の間での事前了解を取ることが望ましい.

#### 2.2 端末内でのパケットの取得

インターネットを介して使用されるアプリケーションの多様性を考慮し、汎用的に通信の監視を行うことのできるパケットキャプチャによって通信内容の取得をするために、

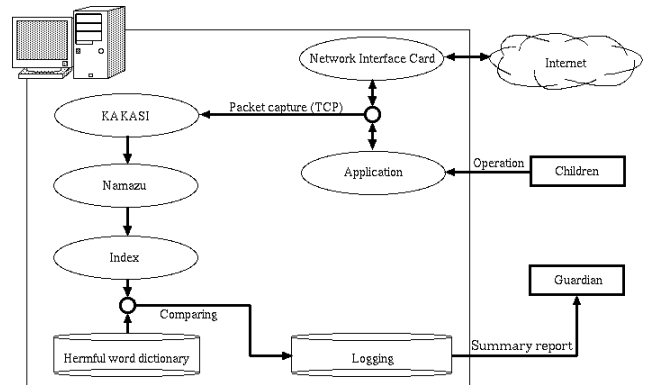


図 1 監視支援システムの概要

Fig. 1 Outline of watchdog system.

パケットドライバ WinPcap を用いたツールを作成した[4]. キャプチャされたパケットは、データリンク層のフレームとして得られ、ヘッダを取り除いた IP データグラムを取り出す. 同様にネットワーク層、トランスポート層のヘッダを参照して TCP/IP の通信のみを選択し、各ヘッダを取り除いた TCP データを得る.

また、これらのヘッダを用いることで、TCP/IP およびアプリケーション層のプロトコルの挙動を模擬する. インターネットとアプリケーション間の通信を再現することにより、通信に含まれる画像、音声などのバイナリデータを取り除き、テキストデータを得ることができる.

本稿では、これらのうち暗号化されていない通信において送受信される Web ドキュメントに含まれる日本語のテキストデータを取得する.

#### 2.3 有害単語およびカテゴリ別単語の辞書化

情報が有害であるかどうかの判断は、受信者によって大きく異なる. 例えば、子どもにとって過激な性表現は有害な情報となるが、そのような情報に興味をもつ成人にとっては有害な情報とならない.

本稿における有害な情報は、子どもにとって有害な情報を意味するものとし、具体的には性、暴力、犯罪、麻薬、人権侵害等を示す. これらの情報を表現するためによく用いられる単語を集めて辞書化する. 表 1 に辞書に含まれる単語の一例を挙げる. また、子どもが興味を持っている分野を調べるために、様々な分野の単語を集めた辞書を作成する.

#### 2.4 単語の検索

Web ドキュメントに含まれるテキストデータに対し分かち書きを行い単語ごとに区切った後、Namazu[5]でインデックスを作成する. インデックスと辞書と比較し、送受信されたパケットにこれらの単語が含まれているかどうかを調べて記録する. これにより、子どもが有害な情報に接触

† 北海道大学大学院情報科学研究科, Graduate school of Information Science and Technology, Hokkaido University

‡ 北海道大学情報基盤センター, Information Initiative Center, Hokkaido University

していないかどうか、どのような情報に興味を持っているかどうかの傾向を知ることができる。

アクセス先の Web アドレスに基づいた検知を行う場合は、有害なサイトのアドレスをまとめたデータベースとの比較により、子どもが有害なサイトにアクセスをしているかどうかの判定を行なうが、開設されて間もないサイトなどデータベースに載っていないサイトへのアクセスを検知することはできない。本システムの手法では、子どもが実際に送受信したパケットに含まれる情報を元に検知を行うため、この問題を回避することができる。

## 2.5 保護者への通知

パケットに含まれる単語を分析した図 2 のような結果の要旨を、あらかじめ登録された保護者の電子メールアドレスに通知する。この要旨には、通信時間、転送量、パケット内の単語の傾向が示される。また、この要旨にはインターネットと PC 間で送受信されたテキストデータが一切含まれていないため、子どものプライバシーを守ることもつながる。単語の辞書を有害なものだけでなく、幅広いカテゴリについて用意することで、保護者が子どもの興味の変遷を知ること可能である。

## 3. システムの動作実験

Windows PC 上に実装されたプロトタイプシステムの動作を確認するために、Impress Watch, Yahoo! Japan News, 2ちゃんねる News カテゴリ, アダルトサイトをそれぞれ 1 時間閲覧し、検出された単語の分布を比較する。

各サイトから受信したドキュメントに含まれる単語の種類および総数は表 2 のようになった。また、単語を辞書と比較して得られた結果を図 3 に示す。2ちゃんねるとアダルトサイトの単語種数が多いが、これはこれらのサイトにおいて口語的表現が多用されており、分かち書きを行うことが難しいことによるものである。

Impress Watch と Yahoo! Japan News では IT に関連する単語、アダルトサイトでは性に関連する単語が多く検出された。また 2ちゃんねるでは複数のカテゴリに渡って有害単語が少量ずつ含まれていることが確認できる。

## 4. まとめ

インターネット利用監視システムの Windows プラットフォーム上への実装と有害単語検出の動作実験について述べた。今後、パソコンにあまり詳しくない保護者に配慮し、簡単に操作できる GUI の実装を目指すとともに、ドキュメントに含まれる単語の傾向を利用することにより、自律的に辞書の学習を行いフィードバックすることで有害判別の精度の向上を図る。

表 1 有害およびカテゴリ辞書に含まれる単語の一例  
Table.1 Example of word included in dictionary.

カテゴリ	単語例
性	アダルトビデオ, イメクラ, ポルノ, 痴漢
ドラッグ	コカイン, シンナー, ヘロイン, 大麻
人権侵害	(放送禁止用語などから抜粋)
2ちゃんねる	きぼんぬ, ぬるぼ, もうだめぼ, インスコ, オマエモナー
IT	イーサネット, エンコード, スパイウェア, 集積回路



図 2 保護者への通知メール  
Fig.2 Notification mail to guardian.

表 2 各サイトを閲覧した結果  
Table.2 Result of browsing each site.

	単語種数	単語総数
Impress Watch	10,259	60,948
Yahoo! Japan News	11,507	124,205
2ちゃんねる	27,900	124,044
アダルトサイト	17,285	120,888

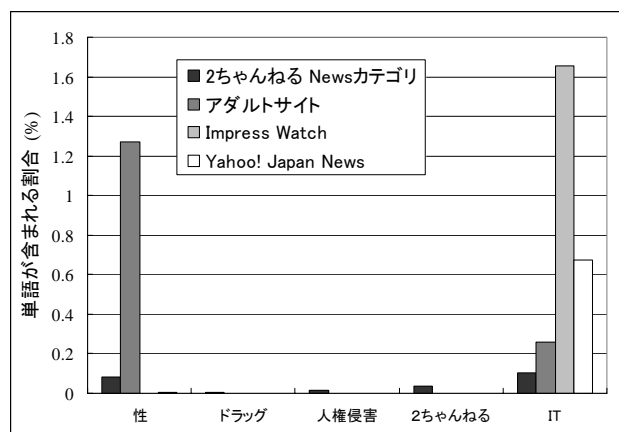


図 3 辞書との比較結果  
Fig.3 Result of comparing with dictionary.

## 参考文献

- [1] 平成 16 年版 情報通信白書, 総務省
- [2] 五味崇, 安田裕樹, “インターネットコンテンツに対する法規制の実現可能性に関する検討”, UFJ Institute REPORT 2002.12 Vol.8 No.1 (2002)
- [3] “フィルタリング情報ページ”, 財団法人インターネット協会, <http://www.iajapan.org/rating/>
- [4] WinPcap, <http://winpcap.polito.it/>
- [5] Namazu, <http://www.namazu.org/>